Upcycling Models under Domain and Category Shift

Sanqing Qu[1*], Tianpei Zou[1*], Florian Röhrbein[2], Cewu Lu[3], Guang Chen[1†], Dacheng Tao[4,5], Changjun Jiang[1]

[1]Tongji University, [2]Chemnitz University of Technology,

[3]Shanghai Jiao Tong University, [4]JD Explore Academy, [5]The University of Sydney

Abstract

Deep neural networks (DNNs) often perform poorly in the presence of domain shift and category shift. How to upcycle DNNs and adapt them to the target task remains an important open problem. Unsupervised Domain Adaptation (UDA), especially recently proposed Source-free Domain Adaptation (SFDA), has become a promising technology to address this issue. Nevertheless, existing SFDA methods require that the source domain and target domain share the same label space, consequently being only applicable to the vanilla closed-set setting. In this paper, we take one step further and explore the Source-free Universal Domain Adaptation (SF-UniDA). The goal is to identify "known" data samples under both domain and category shift, and reject those "unknown" data samples (not present in source classes), with only the knowledge from standard pre-trained source model. To this end, we introduce an innovative global and local clustering learning technique (GLC). Specifically, we design a novel, adaptive one-vs-all global clustering algorithm to achieve the distinction across different target classes and introduce a local k-NN clustering strategy to alleviate negative transfer. We examine the superiority of our GLC on multiple benchmarks with different category shift scenarios, including partial-set, open-set, and open-partial-set DA. Remarkably, in the most challenging open-partial-set DA scenario, GLC outperforms UMAD by 14.8% on the VisDA benchmark. The code is available at https://github.com/ispc-lab/GLC.

1. Introduction

At the expensive cost of given large-scale labeled data and huge computation resources, deep neural networks (DNNs) have made remarkable progress in various tasks. However, DNNs often generalize poorly to the unseen new domain under domain shift and category shift. How to upcycle DNNs and adapt them to target tasks is still a long-standing open problem. In the last decade, many efforts have been devoted to unsupervised domain adaptation (UDA) [12, 16, 28, 40], which capitalizes on labeled source data and unlabeled target data in a transduction manner, and has achieved significant success. Despite this, the access to source raw data is inefficient and may violate the increasingly stringent data privacy policies [45]. Recently, Source-free Domain Adaptation (SFDA) [22, 35, 46] has become a promising technology to alleviate this issue, where only a pre-trained source model is provided as supervision rather than raw data. However, to avoid model collapse, most existing methods [22, 35, 46] assume that the label space is identical across the source and target domain, thus being only applicable to vanilla closed-set scenarios.

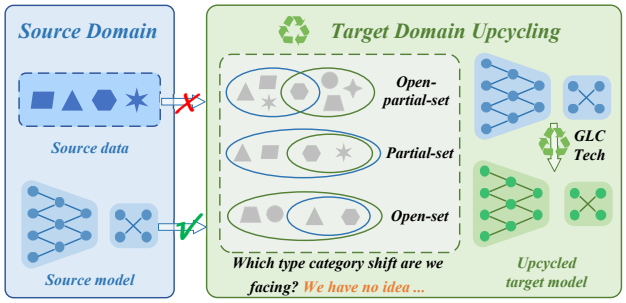


Figure 1. The illustration of Source-free Universal Domain Adaptation (SF-UniDA). The goal is to realize model upcycling under both domain shift and category shift. It is extremely challenging as only one source closed-set model is provided as supervision rather than raw data. And we do not have any prior knowledge about category shift between domains in advance.

In reality, target data may come from a variety of scenarios. Therefore, it is too difficult to hold such a strict assumption. For a better illustration, we suppose $\mathcal{Y}_s$ and $\mathcal{Y}_t$ as the label space of source domain and target domain, respectively. In addition to the well-studied vanilla closed-set ($\mathcal{Y}_s = \mathcal{Y}_t$.), we often encounter several other situations, e.g., the partial-set ($\mathcal{Y}_s \supset \mathcal{Y}_t$), the open-set ($\mathcal{Y}_s \subset \mathcal{Y}_t$), and the open-partial-set ($\mathcal{Y}_s \cap \mathcal{Y}_t \neq \emptyset$, $\mathcal{Y}_s \nsubseteq \mathcal{Y}_t$, $\mathcal{Y}_s \not\supseteq \mathcal{Y}_t$). Currently, there have been several source data-dependent works [4, 5, 26, 32, 38, 47] developed to target category shift. However, methods devised for one situation are commonly

*Equal Contribution

†Corresponding author: guangchen@tongji.edu.cn

infeasible for others. In practice, the target domain is unlabeled and we cannot know which of these category shifts will occur in advance. Not to mention that the requirement to source raw data makes it inefficient and potentially violates data protection policies. To tackle these limitations, and handle those category shifts in a unified manner, in this paper, we take one step further and delve into the Source-free Universal Domain Adaptation (SF-UniDA). The goal is to upcycle the standard pre-trained source models identifying "known" data samples and rejecting those "unknown" data samples (not present in source classes) under domain and category shift. We conceptually present the SF-UniDA in Fig. 1. Note that, very few works [18,23] have studied the source-free model adaptation in open-partial-set scenarios. Nevertheless, their approaches demand dedicated model architectures, greatly limiting their practical applications. SF-UniDA is appealing in view that model adaptation can be resolved only on the basis of a standard pre-trained closed-set model, i.e., without specified model architectures.

To approach such a challenging DA setting, we propose a simple yet generic technique, *Global and Local Clustering (GLC)*. Different from existing pseudo-labeling strategies that focus on closed-set scenarios, we develop a novel one-vs-all global clustering based pseudo-labeling algorithm to achieve "known" data identification and "unknown" data rejection. As we have no prior about the category shift, we utilize the Silhouettes [36] metric to help us realize adaptive global clustering. To avoid source private categories misleading, we design a global confidence statistics based suppression strategy. Although the global clustering algorithm encourages the separation of "known" and "unknown" data samples, we find that some semantically incorrect pseudo-label assignments may still occur, leading to negative knowledge transfer. To mitigate this, we further introduce a local k-NN clustering strategy by exploiting the intrinsic consensus structure of the target domain.

We validate the superiority of our GLC via extensive experiments on four benchmarks (Office-31 [37], Office-Home [44], VisDA [34], and Domain-Net [33]) under various category shift situations, including partial-set, open-set and open-partial-set. Empirical results show that GLC yields state-of-the-art performance across multiple benchmarks, even with stricter constraints.

Our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to exploit and achieve the Source-free Universal Domain Adaptation (SF-UniDA) with only a standard pre-trained closed-set model.
- We propose a generic global and local clustering technique (GLC) to address the SF-UniDA. GLC equips with an innovative global one-vs-all clustering algorithm to realize "known" and "unknown" data samples separation under various category-shift.
- Extensive experiments on four benchmarks under various category-shift situations demonstrate the superiority of our GLC technique. Remarkably, in the open-partial-set DA situation, GLC attains an H-score of 73.1% on the VisDA benchmark, which is 14.8% and 16.7% higher than UMAD and GATE, respectively.

2. Related Work

Unsupervised Domain Adaptation: To alleviate performance degeneration caused by domain shift, unsupervised domain adaptation (UDA) has received considerable interest in recent years. Existing methods can be broadly classified into three categories: discrepancy based, reconstruction based, and adversarial based. Discrepancy based methods [8, 17, 28] usually introduce a divergence criterion to measure the distance between the source and target data distributions, and then achieve model adaptation by minimizing the corresponding criterion. Reconstruction based methods [1, 13, 30] typically introduce an auxiliary image reconstruction task that guides the network to extract domain-invariant features for model adaptation. Inspired by GAN [14], adversarial based approaches [12, 27, 40] leverage domain discriminators to learn domain-invariant features. Despite of effectiveness, these methods typically focus only on the vanilla closed-set domain adaptation.

Universal Domain Adaptation: To handle category-shift, there have been some methods proposed for partial-set [4, 5], open-set [26, 32, 41], and open-partial-set domain adaptation [39, 47]. However, most of these methods are designed for a specified situation, and are typically not applicable to other category-shift situations. As an example, an open-partial-set method [47] even underperforms the source model in the partial-set scenario. Recently, [7, 38] propose a truly universal UDA method, which is applicable to all three category-shift situations. Nevertheless, most existing methods need access to source data during adaptation, which is inefficient and may violate the increasing data protection policies [45].

Source-free Domain Adaptation: Recently, several works [21, 22, 35, 46] have attempted to achieve domain adaptation with knowledge from only the pre-trained source model rather than raw data. However, to avoid model collapse, these methods commonly focus on the vanilla closed-set domain adaptation, significantly limiting their usability. Very recently, few works [18, 23] have studied the source-free domain adaptation in open-partial-set scenarios. Nevertheless, the requirement of dedicated source model architectures, e.g., specialized two-branch open-set recognition frameworks, greatly limits their deployment in reality. In this paper, we target for achieving truly universal model adaptation, including partial-set, open-set, and open-partial-set scenarios, with the knowledge from vanilla source closed-set model.

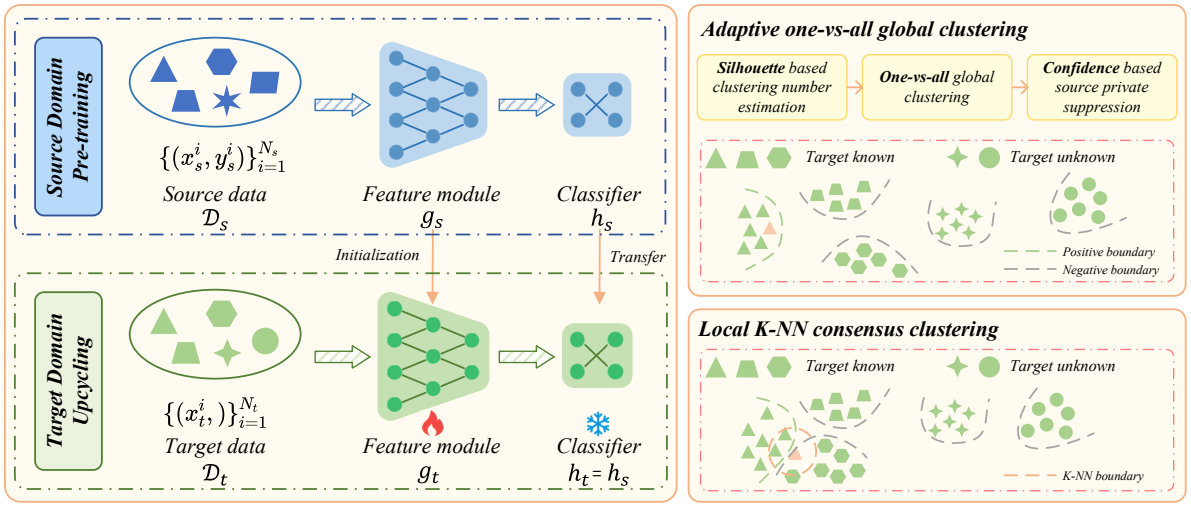


Figure 2. Overview of our proposed Global and Local Clustering technique (GLC). Following the previous source-free closed-set domain adaptation (SFDA) method [22], given a pre-trained source model $f_s = h_s \circ g_s$, we freeze the classifier h_s and merely learn the target-specific feature module g_t by fine-tuning the source feature module g_s for domain alignment. To realize "known" and "unknown" data separation, we develop a novel adaptive one-vs-all global clustering algorithm to assign pseudo labels for each target data sample. As we have no prior about the category shift, we introduce the Silhouette [36] criterion to facilitate us in achieving adaptive one-vs-all clustering. To avoid misleading from source private categories, we develop a global confidence score based suppression strategy. In addition to global clustering, we further exploit the local intrinsic structure to mitigate negative transfer. (Best view in color.)

3. Methodology

3.1. Preliminary

In this paper, we aim to achieve model upcycling under both domain shift and category shift, i.e., the source-free universal domain adaptation (SF-UniDA). In particular, we consider the K-way classification. In this setting, there is a well-designed source domain $\mathcal{D}_s = \{(x_s^i, y_s^i)\}_{i=1}^{N_s}$ where $x_s^i \in \mathcal{X}_s$, $y_s^i \in \mathcal{Y}_s$, and an unlabeled target domain $\mathcal{D}_t = \{(x_t^i, ?)\}_{i=1}^{N_t}$ where $x_t^i \in \mathcal{X}_t$. For a better illustration, we denote $\mathcal{Y} = \mathcal{Y}_s \cap \mathcal{Y}_t$ as the common label space, $\bar{\mathcal{Y}}_s = \mathcal{Y} \backslash \mathcal{Y}_t$ as the source private label space, and $\bar{\mathcal{Y}}_t = \mathcal{Y} \setminus \mathcal{Y}_s$ as the target private label space, respectively. As aforementioned, there are three possible category shifts, i.e., the partial-set DA, PDA, ($\mathcal{Y}_s \supset \mathcal{Y}_t$); the open-set DA, OSDA, ($\mathcal{Y}_s \subset \mathcal{Y}_t$); and the open-partial-set DA, OPDA, ($\mathcal{Y} \neq \emptyset$, $\bar{\mathcal{Y}}_s \neq \emptyset$, $\bar{\mathcal{Y}}_t \neq \emptyset$). The final goal is to identify "known" samples (belonging to $\mathcal{Y}$) and reject "unknown" samples (belonging to $\bar{\mathcal{Y}}_t$) of $\mathcal{D}_t$, with the knowledge only from source pre-trained model f_s. $\mathcal{D}_s$ is not available, and we do not have prior knowledge of what kind of category shift we are facing.

There have been few works [18, 23] explored source-free domain adaptation under category shift. However, these methods are limited to specific category shift, and require dedicated source model architectures. To address these limitations, we propose to achieve SF-UniDA on the basis of only the vanilla closed-set model. Following existing closed-set source-free domain adaptation methods [22, 24], given a source model $f_s = h_s \circ g_s$, consisting of a feature module g_s and a classifier module h_s, we capitalize on the source hypothesis to achieve source and target domain alignment. That is, we only learn a target-specific feature module g_t and keep the classifier $h_t = h_s$. To realize "known" data identification and "unknown" data rejection under both domain shift and category shift, we devise a novel, adaptive, global one-vs-all clustering algorithm. Besides, we further employ a local k-NN clustering strategy to alleviate negative transfer. The pipeline is presented in Fig. 2. More details will be described in the following.

3.2. One-vs-all Global Clustering

Pseudo-labeling is a promising technique in unsupervised learning. Traditional pseudo-labeling strategies [19] assign pseudo labels directly based on sample-level predictions, which are often noisy, especially in the presence of domain shifts. To mitigate this, there are some pseudo-labeling strategies [22, 31] exploit the data structure of the target domain, *i.e.*, the target-specific prototypes. However, these strategies assume that the source and target domain share identical label space, making it infeasible under category shift. Therefore, a question naturally arises: *How to achieve pseudo-labeling with inconsistent label space? Especially, for universal domain adaptation, we have no prior about the category shift between $\mathcal{Y}_s$ and $\mathcal{Y}_t$.*

To tackle this, we first view this problem from a simplified perspective: *If $\mathcal{Y}_s \subset \mathcal{Y}_t$ (i.e., the OSDA setting), and we were to know the number of categories in the target domain is C_t, what kind of pseudo-labeling strategy should we apply?* Intuitively, target domain, in this case, should be grouped into C_t clusters, each corresponding to a specific

category. We can then assign pseudo labels via the nearest cluster centroid classifier. However, even though we apply existing clustering algorithms, such as K-means [29], to divide the target domain into C_t clusters. It is still challenging to associate the corresponding semantic category for each cluster, in particular for the SF-UniDA, as we have no access to the source raw data.

In view of this, to ease the challenging semantic association, we devise a novel one-vs-all global clustering pseudo-labeling algorithm. The main idea is that *For a particular "known" category $c \in C_s$, in order to decide whether a data sample belongs to the c-th category, we need to figure out what is and what is not the c-th category.* The detailed procedure is presented as follows:

- For a particular c-th category, we first aggregate the top-K $\delta_c(f_t(x_t))$ scores represented instances along all target domain $\mathcal{D}_t$ as positive $\mathcal{P}_c$, and the rest as negative $\mathcal{N}_c$. Here, $\delta_c(f_t(x_t))$ denotes the soft-max probability of target instance x_t belonging to the c-th class. We empirically set $K = N_t/C_t$.

- Then, we obtain the positive prototype representation p_c (*i.e.*, what is the c-th category), and negative prototypes $\{n_c^i\}_{i=1}^M$ (*i.e.*, what are not the c-th category) via K-means. Noting that we have employed multiple prototypes to represent the negatives since the negatives contain distinct classes. We set M to C_t instead of $C_t - 1$, considering that the "known" category of the target domain typically involves some hard samples that are difficult to be selected by top-K sampling.

$$\begin{aligned} p_c &= \frac{1}{K} \sum_{x_t \in \mathcal{P}_c} g_t(x_t), \\ \{n_c^i\}_{i=1}^M &= \underset{x_t \in \mathcal{N}_c}{Kmeans}(g_t(x_t)). \end{aligned} \tag{1}$$

- Thereafter, we decide whether data sample x_t belongs to the c-th category via the nearest centroid classifier:

$$\hat{q}_c = \begin{cases} 1, \text{if } S(g_t(x_t), p_c) \geq \max\{S(g_t(x_t), n_c^i)\}_{i=1}^M \\ 0, \text{if } S(g_t(x_t), p_c) < \max\{S(g_t(x_t), n_c^i)\}_{i=1}^M \end{cases} \tag{2}$$

  where $S(a, b)$ measures the similarity between a and b. We apply the cosine similarity function by default.

- Finally, we iterate the above process to obtain the pseudo labels $\hat{y}_t$ for all "known" category $c \in C_s$. Since each data sample either belongs to the unknown or to one of the categories in the source domain, it is not possible to belong to multiple categories at the same time. Thereby, we introduce a filtering strategy to avoid semantic ambiguity. Here, we just set the category with maximum similarity as the target. It is worth noting that our algorithm does not require the above pseudo-label $\hat{y}_t$ to be one-hot encoded. Those pseudo labels with all-zero encoding mean that these data samples belong to the "unknown" target-private categories $\bar{\mathcal{Y}}_t$. To realize "known" and "unknown" separation, we then manually set those all-zero encoding pseudo labels to a uniform encoding, *i.e.*, $\hat{q}_c = 1/C_s$.

3.3. Confidence based Source-private Suppression

In the above section, we developed the one-vs-all global clustering algorithm to assign pseudo labels for OSDA, *i.e.*, $\mathcal{Y}_s \subset \mathcal{Y}_t$, when the number of categories in the target domain C_t is available. However, in addition to OSDA, we may also encounter PDA and OPDA, where the source domain contains categories absent in the target domain. To make the above algorithm applicable to both OSDA, PDA and OPDA, it is necessary to tailor the proposed algorithm to prevent those source-private categories from misleading pseudo-label assignments.

We empirically found that on positive data group $\mathcal{P}$ sampled with top-K on the target domain, those source-private categories generally yield lower mean prediction confidence than those source-target shared categories. In light of this observation, we design a source-private category suppression strategy based on the mean prediction confidence of the positive data group $\mathcal{P}$. Specifically, for a particular category $c \in C_s$, we tailored the Eq. 2 to:

$$\begin{aligned} \epsilon_c &= \rho + \frac{1-\rho}{K} \sum_{x_t \in \mathcal{P}_c} \delta_c(f_t(x_t)), \\ \hat{q}_c &= \begin{cases} 1, \text{if } \epsilon_c \cdot S(g_t(x_t), p_c) \geq \max\{S(g_t(x_t), n_c^i)\}_{i=1}^M \\ 0, \text{if } \epsilon_c \cdot S(g_t(x_t), p_c) < \max\{S(g_t(x_t), n_c^i)\}_{i=1}^M \end{cases} \end{aligned} \tag{3}$$

where ϵ_c is the designed source-private suppression weight for the c-th category, and ρ is a hyper-parameter to control this weight. We empirically set ρ to 0.75 for all datasets. Its sensitivity analysis can be found in the experiment.

3.4. Silhouette Based Target Domain C_t Estimation

Based on the previous sections, we now have achieved the pseudo-labeling algorithm for SF-UniDA. However, it is still not applicable yet, due to the requirement of prior information, i.e., the number of categories C_t in the target domain, which is commonly unavailable in reality. Therefore, the last obstacle for us is: *How to determine the number of categories C_t in the target domain?*

To address this, a feasible solution is to first enumerate the possible values of the number of categories C_t in the target domain and divide the target domain into the corresponding clusters by applying a clustering algorithm like K-means [29]. Then the clustering evaluation criteria [3, 9, 36, 43] can be employed to determine the appropriate number of target domain categories $\tilde{C}_t$.

In this paper, we employ the Silhouette criterion [36] to facilitate estimating $\tilde{C}_t$. Technically, for a data sample $x_t \in \mathcal{C}_I$, the Silhouette value $s(x_t)$ is defined as:

$$\begin{aligned} a(x_t) &= \frac{1}{|\mathcal{C}_I| - 1} \sum_{x \in \mathcal{C}_I, x \neq x_t} d(x_t, x), \\ b(x_t) &= \min_{J \neq I} \frac{1}{|\mathcal{C}_J|} \sum_{x \in \mathcal{C}_J} d(x_t, x), \\ s(x_t) &= \frac{b(x_t) - a(x_t)}{\max\{a(x_t), b(x_t)\}}. \end{aligned} \tag{4}$$

where $a(x_t)$ and $b(x_t)$ measure the similarity of x_t to its own cluster $\mathcal{C}_I$ (cohesion) and other clusters $\mathcal{C}_{J,J\neq I}$ (separation), respectively. $d(x_i, x_j)$ measures the distance between data points x_i and x_j, and $|\mathcal{C}_I|$ denotes the size of cluster $\mathcal{C}_I$. The Silhouette value $s(x_t)$ ranges from -1 to +1, where a high value indicates that the data sample x_t has a high match with its own cluster and a low match with neighboring clusters. Therefore, if most of the data samples have high Silhouette values, then the clustering configuration is appropriate; otherwise, the clustering configuration may have too many or too few clusters.

Since it is challenging to obtain the exact number of target domain categories C_t, in our implementation, we empirically enumerate the possible values of $\tilde{C}_t$ as $[1/3C_s, 1/2C_s, C_s, 2C_s, 3C_s]$, taking into account the scenarios may encounter. Note that we only estimate the value of $\tilde{C}_t$ at the beginning, and subsequently, we do not change the value of $\tilde{C}_t$ considering the overall efficiency.

3.5. Local Consensus Clustering

Although the global one-vs-all clustering pseudo-labeling algorithm encourages the separation between "known" and "unknown" data samples, semantically incorrect pseudo-label assignments still occur due to domain shift and category shift, resulting in negative transfer.

To mitigate this, we further introduce a local k-NN consensus clustering strategy that exploits the intrinsic consensus structure of the target domain $\mathcal{D}_t$. Specifically, during model adaptation, we maintain a memory bank $\mathcal{G}_t = \{g_t(x_t), \delta(f_t(x_t))\}_{x_t \in \mathcal{D}_t}$, which contains the target features and corresponding prediction scores. The local k-NN consensus clustering is then realized by:

$$\begin{aligned} l_c^i &= \frac{1}{|L^i|} \sum_{x_t \in L^i} \delta_c(f_t(x_t)), \\ \mathcal{L}_{tar}^{loc} &= -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C_s} l_c^i \log \delta_c(f_t(x_t^i)). \end{aligned} \tag{5}$$

where $\delta_c(f_t(x_t))$ denotes the soft-max probability of data instance x_t belonging to the c-th class, L^i refers to the set of nearest neighbors of data x_t^i in the embedding feature space. Here, we apply the cosine similarity function to find the nearest neighbors L^i of x_t^i in the memory bank $\mathcal{G}_t$. We then encourage minimizing the cross entropy loss between x_t^i and the nearest neighbors L^i to achieve the local semantic consensus clustering.

3.6. Optimization Objective

The overall training loss of GLC can be written as:

$$\begin{aligned} \mathcal{L}_{tar}^{glb} &= -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C_s} \hat{q}_c^i \log \delta_c(h_t(g_t(x_t^i)))), \\ \mathcal{L}_{tar} &= \eta \mathcal{L}_{tar}^{glb} + \mathcal{L}_{tar}^{loc}. \end{aligned} \tag{6}$$

where $\hat{q}_c^i$ denotes the global clustering pseudo label for data sample x_t^i, and $\mathcal{L}_{tar}^{glb}$ is the corresponding global cross-entropy loss. $\eta > 0$ is a trade-off hyper-parameter.

3.7. Inference Details

As there is only one standard classification model, we apply the normalized Shannon Entropy [42] as the uncertainty metric to separate known and unknown data samples:

$$I(x_t) = -\frac{1}{\log C_s} \sum_{c=1}^{C_s} \delta_c(f_t(x_t)) \log \delta_c(f_t(x_t)) \tag{7}$$

where C_s is the class number of source domain $\mathcal{D}_s$, and $\delta_c(f_t(x_t))$ denotes the soft-max probability of data sample x_t belonging to the c-th class. The higher the uncertainty, the more the model f_t tends to assign an unknown label to the data sample. During inference stage, given an input sample x_t, we first compute $I(x_t)$ and then predict the class of $y(x_t)$ with a pre-defined threshold ω as:

$$y(x_t) = \begin{cases} \text{unknown}, & \text{if } I(x_t) \geq \omega \\ \text{argmax}(f_t(x_t)), & \text{if } I(x_t) < \omega \end{cases} \tag{8}$$

which either rejects the input sample x_t as unknown or classifies it into a known class. In our implementation, we set $\omega = 0.55$ for all standard benchmark datasets. Its sensitivity analysis can be found in the experiments.

4. Experiments

4.1. Setup

Dataset: We utilize the following standard datasets in DA to evaluate the effectiveness and versatility of our method. **Office-31** [37] is a widely-used small-sized domain adaptation benchmark, consisting of 31 object classes (4,652 images) under office environment from three domains (DSLR (D), Amazon (A), and Webcam (W)). **Office-Home** [44] is another popular medium-sized benchmark, consisting of 65 categories (15,500 images) from four domains (Artistic images (Ar), Clip-Art images (Cl), Product

Table 1. H-score (%) comparison in OPDA scenario on Office-Home.Some results are cited from GATE [7] and UMAD [23]. SF denotes source data-free. We compare GLC with SF methods and non-SF methods. (Best in red and second best in blue)

Methods	SF	OPDA	OSDA	PDA	Ar2Cl	Ar2Pr	Ar2Re	Cl2Ar	Cl2Pr	Cl2Re	Pr2Ar	Pr2Cl	Pr2Re	Re2Ar	Re2Cl	Re2Pr	Avg
UAN [47]	✗	✓	✗	✗	51.6	51.7	54.3	61.7	57.6	61.9	50.4	47.6	61.5	62.9	52.6	65.2	56.6
CMU [11]	✗	✓	✗	✗	56.0	56.9	59.2	67.0	64.3	67.8	54.7	51.1	66.4	68.2	57.9	69.7	61.6
DCC [20]	✗	✓	✓	✓	58.0	54.1	58.0	**74.6**	70.6	77.5	64.3	**73.6**	74.9	**81.0**	**75.1**	80.4	70.2
OVANet [39]	✗	✓	✓	✗	62.8	75.6	78.6	70.7	68.8	75.0	71.3	58.6	80.5	76.1	64.1	78.9	71.8
GATE [7]	✗	✓	✓	✓	**63.8**	75.9	81.4	**74.0**	**72.1**	**79.8**	**74.7**	**70.3**	**82.7**	79.1	**71.5**	**81.7**	**75.6**
Source-only	✓	-	-	-	47.3	71.6	81.9	51.5	57.2	69.4	56.0	40.3	76.6	61.4	44.2	73.5	60.9
SHOT-O [22]	✓	✗	✓	✗	32.9	29.5	39.6	56.8	30.1	41.1	54.9	35.4	42.3	58.5	33.5	33.3	40.7
UMAD [23]	✓	✓	✓	✗	61.1	**76.3**	**82.7**	70.7	67.7	75.7	64.4	55.7	76.3	73.2	60.4	77.2	70.1
GLC	✓	✓	✓	✓	**64.3**	**78.2**	**89.8**	63.1	**81.7**	**89.1**	**77.6**	54.2	**88.9**	**80.7**	54.2	**85.9**	**75.6**

Table 2. H-score (%) comparison in OPDA scenario on Office-31, VisDA, and DomainNet. Some results are cited from UMAD [23].

Methods	SF	OPDA	OSDA	PDA	Office-31							VisDA	DomainNet						
					A2D	A2W	D2A	D2W	W2A	W2D	Avg	S2R	P2R	P2S	R2P	R2S	S2P	S2R	Avg
UAN [47]	✗	✓	✗	✗	59.7	58.6	60.1	70.6	60.3	71.4	63.5	34.8	41.9	39.1	43.6	38.7	38.9	43.7	41.0
CMU [11]	✗	✓	✗	✗	68.1	67.3	71.4	79.3	72.2	80.4	73.1	32.9	50.8	45.1	52.2	45.6	44.8	51.0	48.3
DCC [20]	✗	✓	✓	✓	**88.5**	78.5	70.2	79.3	75.9	88.6	80.2	43.0	56.9	43.7	50.3	43.3	44.9	56.2	49.2
OVANet [39]	✗	✓	✓	✗	85.8	79.4	80.1	**95.4**	84.0	**94.3**	86.5	53.1	56.0	47.1	51.7	44.9	47.4	**57.2**	50.7
GATE [7]	✗	✓	✓	✓	**87.7**	**81.6**	84.2	**94.8**	83.4	94.1	**87.6**	56.4	57.4	**48.7**	**52.8**	**47.6**	**49.5**	56.3	**52.1**
Source-only	✓	-	-	-	70.9	63.2	39.6	77.3	52.2	86.4	64.9	25.7	57.3	38.2	47.8	38.4	32.2	48.2	43.7
SHOT-O [22]	✓	✗	✓	✗	73.5	67.2	59.3	88.3	77.1	84.4	75.0	44.0	35.0	30.8	37.2	28.3	31.9	32.2	32.6
UMAD [23]	✓	✓	✓	✗	79.1	77.4	**87.4**	90.7	**90.4**	**97.2**	87.0	**58.3**	**59.0**	44.3	50.1	42.1	32.0	55.3	47.1
GLC	✓	✓	✓	✓	81.5	**84.5**	**89.8**	90.4	**88.4**	92.3	**87.8**	**73.1**	**63.3**	**50.5**	**54.9**	**50.9**	**49.6**	**61.3**	**55.1**

Table 3. Details of class split. Here, $\mathcal{Y}$, $\bar{\mathcal{Y}}_s$, and $\bar{\mathcal{Y}}_t$ denotes the source-target-shared class, the source-private class, and the target-private class, respectively.

Dataset	Class Split($\mathcal{Y}/\bar{\mathcal{Y}}_s/\bar{\mathcal{Y}}_t$)		
	OPDA	OSDA	PDA
Office-31 [37]	10/10/11	10/0/11	10/21/0
Office-Home [44]	10/5/50	25/0/40	25/40/0
VisDA-C [34]	6/3/3	6/0/6	6/6/0
DomainNet [33]	150/50/145	-	-

images (Pr), and Real-World images (Rw)). **VisDA-C** [34] is a more challenging benchmark with 12 object classes, where the source domain contains 152,397 synthetic images generated by rendering 3D models and the target domain consists of 55,388 images from Microsoft COCO. **DomainNet** [33], is the largest domain adaptation benchmark with about 0.6 million images, which contains 345 classes. Similar to previous works [7, 23], we conduct experiments on three subsets from it (Painting (P), Real (R), and Sketch (S)). We evaluate our GLC on partial-set DA (PDA), open-set DA (OSDA), and open-partial-set DA (OPDA) scenarios. Detailed classes split are summarized in Table 3.

Evaluation protocols: For a fair comparison, we utilize the same evaluation metric as previous works [7, 20]. Specifically, in PDA scenario, we report the classification accuracy over all target samples. In OSDA and OPDA scenarios, considering the trade-off between "known" and "unknown" categories, we report the H-score, i.e., the harmonic mean of the accuracy of "known" and "unknown" samples.

Implementation details: We adopt the same network architecture with existing baseline methods. Specifically, we adopt the **ResNet-50** [15] pre-trained on ImageNet [10] as the backbone for all datasets. For preparing the source model, here, we utilize the same network structure and training recipe as SHOT [22]. We present more details about source model training in the supplementary. During target model adaptation, we apply the SGD optimizer with momentum 0.9. The batch size is set to 64 for all benchmark datasets. We set the learning rate to 1e-3 for Office-31 and Office-Home, and 1e-4 for VisDA and DomainNet. For hyper-parameter, as we described in previous sections, we set ρ to 0.75 for all datasets. For local k-NN consensus clustering, $|L|$ is set to 4 for all benchmarks. As for η, we set it to 0.3 for Office-31, VisDA, and 1.5 for Office-Home and DomainNet. All experiments are conducted on an RTX-3090 GPU with PyTorch-1.10.

4.2. Experiment Results

To verify the effectiveness of our GLC, we conduct extensive experiments on three possible category-shift scenarios, i.e., open-partial-set DA (OPDA), open-set DA (OSDA), and partial-set DA (PDA). We compare GLC with data-dependent and more recent data-free methods to empirically demonstrate the merit of GLC. In adaptation, data-

Table 4. H-score (%) comparison in OSDA scenario on Office-Home, Office-31, and VisDA. (Best in red and second best in blue)

Methods	SF	OPDA	OSDA	PDA	Office-Home													Office31	VisDA
					Ar2Cl	Ar2Pr	Ar2Re	Cl2Ar	Cl2Pr	Cl2Re	Pr2Ar	Pr2Cl	Pr2Re	Re2Ar	Re2Cl	Re2Pr	Avg	Avg	Avg
OSBP [41]	✗	✗	✓	✗	55.1	65.2	72.9	**64.3**	64.7	70.6	63.2	53.2	73.9	66.7	54.5	72.3	64.7	83.7	52.3
ROS [2]	✗	✗	✓	✗	60.1	69.3	76.5	58.9	65.2	68.6	60.6	56.3	74.4	**68.8**	60.4	**75.7**	66.2	85.9	66.5
CMU [11]	✗	✓	✗	✗	55.0	57.0	59.0	59.3	58.2	60.6	59.2	51.3	61.2	61.9	53.5	55.3	57.6	65.2	54.2
DANCE [38]	✗	✓	✓	✓	6.5	9.0	9.9	20.4	10.4	9.2	28.4	12.8	12.6	14.2	7.9	13.2	12.9	79.8	67.5
DCC [20]	✗	✓	✓	✓	56.1	67.5	66.7	49.6	66.5	64.0	55.8	53.0	70.5	61.6	57.2	71.9	61.7	72.7	59.6
OVANet [39]	✗	✓	✓	✗	58.6	66.3	69.9	62.0	65.2	68.6	59.8	53.4	69.3	68.7	59.6	66.7	64.0	**91.7**	66.1
GATE [7]	✗	✓	✓	✓	**63.8**	70.5	75.8	**66.4**	67.9	71.7	**67.3**	**61.5**	**76.0**	**70.4**	**61.8**	75.1	**69.0**	89.5	**70.8**
Source-only	✓	-	-	-	48.7	65.2	72.6	47.0	56.8	61.8	50.6	39.1	67.8	62.9	48.1	68.4	57.4	73.0	33.1
SHOT-O [22]	✓	✗	✓	✗	37.7	41.8	48.4	56.4	39.8	40.9	60.0	41.5	49.7	61.8	41.4	43.6	46.9	77.5	28.1
UMAD [23]	✓	✓	✓	✗	59.2	**71.8**	**76.6**	63.5	**69.0**	**71.9**	62.5	54.6	72.8	66.5	57.9	70.7	66.4	**89.8**	66.8
GLC	✓	✓	✓	✓	**65.3**	**74.2**	**79.0**	60.4	**71.6**	**74.7**	**63.7**	**63.2**	**75.8**	67.1	**64.3**	**77.8**	**69.8**	89.0	**72.5**

Table 5. Accuracy (%) comparison in PDA scenario on Office-Home, Office-31, and VisDA. (Best in red and second best in blue)

Methods	SF	OPDA	OSDA	PDA	OfficeHome													Office31	VisDA
					Ar2Cl	Ar2Pr	Ar2Re	Cl2Ar	Cl2Pr	Cl2Re	Pr2Ar	Pr2Cl	Pr2Re	Re2Ar	Re2Cl	Re2Pr	Avg	Avg	Avg
ETN [6]	✗	✗	✗	✓	59.2	77.0	79.5	62.9	65.7	75.0	68.3	55.4	84.4	75.7	57.7	84.5	70.4	**96.7**	59.8
BA3US [25]	✗	✗	✗	✓	**60.6**	**83.2**	**88.4**	71.8	**72.8**	83.4	**75.5**	61.6	**86.5**	79.3	62.8	**86.1**	**76.0**	**97.8**	54.9
DANCE [38]	✗	✓	✓	✓	53.6	73.2	84.9	70.8	67.3	82.6	70.0	50.9	84.8	77.0	55.9	81.8	71.1	86.0	73.7
DCC [20]	✗	✓	✓	✓	54.2	47.5	57.5	**83.8**	71.6	**86.2**	63.7	**65.0**	75.2	**85.5**	**78.2**	82.6	70.9	93.3	72.4
OVANet [39]	✗	✓	✓	✗	34.1	54.6	72.1	42.4	47.3	55.9	38.2	26.2	61.7	56.7	35.8	68.9	49.5	74.6	34.3
GATE [7]	✗	✓	✓	✓	55.8	75.9	85.3	73.6	70.2	83.0	72.1	59.5	84.7	79.6	**63.9**	83.8	74.0	93.7	**75.6**
Source-only	✓	-	-	-	45.9	69.2	81.1	55.7	61.2	64.8	60.7	41.1	75.8	70.5	49.9	78.4	62.9	87.8	42.8
SHOT-P [22]	✓	✗	✗	✓	**64.7**	**85.1**	**90.1**	**75.1**	**73.9**	**84.2**	**76.4**	**64.1**	**90.3**	**80.7**	63.3	**85.5**	**77.8**	92.2	74.2
UMAD [23]	✓	✓	✓	✗	51.2	66.5	79.2	63.1	62.9	68.2	63.3	56.4	75.9	74.5	55.9	78.3	66.3	89.5	68.5
GLC	✓	✓	✓	✓	55.9	79.0	87.5	72.5	71.8	82.7	74.9	41.7	82.4	77.3	60.4	84.3	72.5	94.1	**76.2**

dependent methods typically require access to source raw-data, while data-free methods require source pre-trained models. In particular, GLC requires only a standard pre-trained source model, i.e., without any dedicated model architectures as [18, 23]. For a fair comparison, all methods are performed without the prior knowledge of category-shift, except those designed only for specific scenarios.

Results on OPDA: We first conduct experiments on the most challenging setting, i.e., OPDA, in which both source and target domains involve private categories. Results on Office-Home are summarized in Table 1, and results on Office-31, VisDA and DomainNet are summarized in Table 2. As shown in Table 1 and Table 2, our GLC achieves new state-of-the-arts, even compared to previous data-dependent methods. Especially, on VisDA, GLC achieves the H-score of 73.1%, which surpasses GATE [7] and UMAD [23] by a wide margin (16.7% and 14.8%). On the largest benchmark, i.e., DomainNet, GLC still achieves consistent performance improvements compared to UMAD and GATE, with gains of approximately 8.0% and 3.0%.

Results on OSDA: We then conduct experiments on OSDA, where only the target domain involves categories not presented in the source domain. Results on Office-Home, Office-31, and VisDA are summarized in Table 4. As shown in Table 4, GLC still achieves state-of-the-art performance. Specifically, GLC obtains 69.8% H-score on Office-Home and 72.5% H-score on VisDA, with an improvement of 3.4% and 5.7% compared to UMAD.

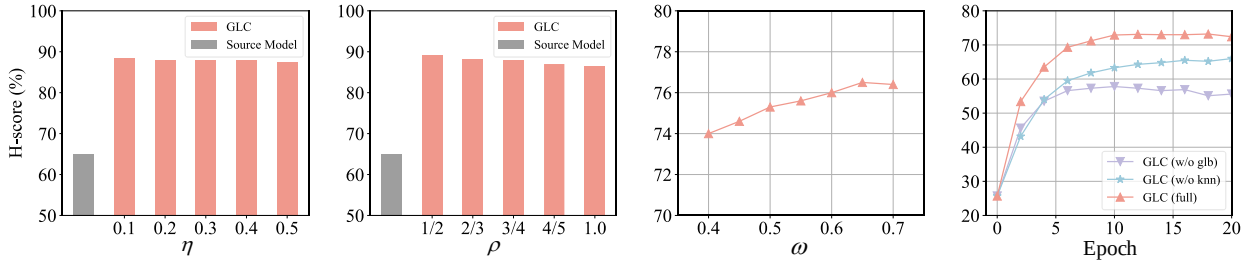
Table 6. **Ablation Study**. Results for OPDA on Office-31, Office-Home, and VisDA with different variants of GLC.

Method	Office-31	Office-Home	VisDA
Source model	64.9	60.9	25.7
GLC (w/o $\mathcal{L}_{tar}^{loc}$)	86.1	74.8	66.0
GLC (w/o $\mathcal{L}_{tar}^{glb}$)	87.4	67.2	57.3
GLC (full)	**87.8**	**75.6**	**73.1**

Results on PDA: We last verify the effectiveness of GLC on PDA, where the label space of the target domain is a subset of the source domain. Results summarized in Table 5 show that GLC still achieves comparable performance compared to methods tailored for PDA. In a fairer comparison, GLC clearly outperforms UMAD, specifically achieving performance gains of 6.2%, 4.6%, and 7.7% on Office-31, Office-Home, and VisDA, respectively.

4.3. Experiment Analysis

Ablation Study: To verify the effectiveness of different components within GLC, we conduct extensive ablation studies on Office-31, Office-Home, and VisDA in OPDA scenarios. The results are summarized in Table 6. Here *GLC* w/o $\mathcal{L}_{tar}^{glb}$ refers to that we only employ the local k-NN consensus clustering loss to regulate model adapta-



(a) sensitivity to η ($\rho = 3/4$) (b) sensitivity to ρ ($\eta = 0.3$) (c) H-score w.r.t ω (d) H-score convergence

Figure 3. **Analysis of GLC.** (a-b) present the hyper-parameter sensitivity of η and ρ on Office-31 in OPDA. (c) plots the H-score with respect to ω on Office-Home in OPDA. (d) shows the H-score curves on VisDA in OPDA during the training process. Here GLC (w/o glb) refers to *GLC* w/o $\mathcal{L}_{tar}^{glb}$ and GLC w/o knn denotes *GLC w/o* $\mathcal{L}_{tar}^{loc}$.

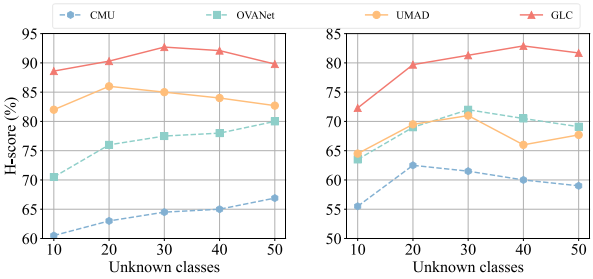
tion, while *GLC* w/o $\mathcal{L}_{tar}^{loc}$ denotes that we only employ the global one-vs-all clustering based pseudo-labeling algorithm to achieve model adaptation. From these results, we can conclude that our local and global clustering strategies are complementary to each other. And global clustering is of vital importance to help us distinguish "known" and "unknown" categories. For example, on VisDA, with only $\mathcal{L}_{tar}^{glb}$, we can advance the source model from the H-score of 25.7% to 66.0%, and outperform GATE by 9.6%.

Hyper-parameter Sensitivity: We first study the parameter sensitivity of η and ρ on Office-31 under OPDA setting in Fig. 3 (a-b), where η is in the range of [0.1, 0.2, 0.3, 0.4, 0.5], and ρ is in the range of [1/2, 2/3, 3/4, 4/5, 1.0]. Note that $\rho = 1.0$ denotes that we do not introduce the confidence based source-private suppression mechanism. It is easy to find that results around the selected parameters $\eta = 0.3$ and $\rho = 0.75$ are stable, and much better than the source model. By oracle validation, we may find better hyper-parameter settings, e.g., $\eta = 0.1$ and $\rho = 0.50$. In Fig. 3 (c), we present the H-score with respect to ω on Office-Home in OPDA. For all benchmarks, we pre-define $\omega = 0.55$ to separate "known" and "unknown" samples. The results show that H-score is relatively stable around our selection, and we could achieve better performance when setting ω to 0.65 via oracle validation. Besides, in Fig. 3 (d), we illustrate the H-score convergence curves on VisDA.

Varying Unknown Classes: As increasing "unknown" classes, it becomes more difficult to correctly identify the "unknown" and "known" objects. To examine the robustness of GLC, we compare GLC with other methods when varying unknown classes on Office-Home under OPDA setting. Fig. 4 shows that GLC achieves more stable and much better performance against existing methods.

4.4. Discussion

So far, most existing domain adaptation methods designed for category shift are not applicable to the vanilla closed-set DA (CLDA). To verify the effectiveness of GLC in CLDA, we have conducted experiments on Office-31 and Office-Home in the Appendix. Moreover, existing methods usually perform experiments only on standard computer



(a) Ar →Re in OPDA (b) Cl → Pr in OPDA

Figure 4. **H-score (%) of OPDA** when varying the number of unknown classes in Office-Home. GLC shows stable and much superior performance against existing methods.

science benchmarks. Here, we have validated the effectiveness of GLC in more realistic applications, including remote sensing in PDA, wildlife classification in OSDA, and single-cell RNA sequence identification in OPDA. These results are also presented in the Appendix.

5. Conclusion

In this paper, we have presented *Global and Local Clustering (CLC)* for upcycling models under domain shift and category shift. Technically, we have devised an innovative one-vs-all global clustering strategy to realize "unknown" and "known" data separation, and introduced a local k-NN clustering strategy to alleviate negative transfer. Compared to existing approaches that require source data or are only applicable to specific category shifts, GLC is appealing by enabling universal model adaptation on the basis of only standard pre-trained source models. Extensive experiments in partial-set, open-set, and open-partial-set DA scenarios across several benchmarks have verified the effectiveness and superiority of GLC. Remarkably, GLC significantly outperforms existing methods by almost 15% on the VisDA benchmark in open-partial-set DA scenario.

Acknowledgment: This work is supported by Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), ZJ Lab, and Shanghai Center for Brain Science and Brain-Inspired Technology, and the Shanghai Rising Star Program (No.21QC1400900).

References

[1] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *NeurIPS*, 2016. 2

[2] Silvia Bucci, Mohammad Reza Loghmani, and Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *ECCV*, 2020. 7

[3] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974. 4

[4] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Partial transfer learning with selective adversarial networks. In *CVPR*, 2018. 1, 2

[5] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *ECCV*, 2018. 1, 2

[6] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. In *CVPR*, 2019. 7

[7] Liang Chen, Yihang Lou, Jianzhong He, Tao Bai, and Minghua Deng. Geometric anchor correspondence mining with uncertainty modeling for universal domain adaptation. In *CVPR*, 2022. 2, 6, 7

[8] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE TPMAI*, 2016. 2

[9] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE TPAMI*, 1979. 4

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 6

[11] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *ECCV*, 2020. 6, 7

[12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *JMLR*, 2016. 1, 2

[13] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *ECCV*, 2016. 2

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 2

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 6

[16] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *ICML*, 2018. 1

[17] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *CVPR*, 2019. 2

[18] Jogendra Nath Kundu, Naveen Venkat, R Venkatesh Babu, et al. Universal source-free domain adaptation. In *CVPR*, 2020. 2, 3, 7

[19] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML Worshop*, 2013. 3

[20] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. In *CVPR*, 2021. 6, 7

[21] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *CVPR*, 2020. 2

[22] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*, 2020. 1, 2, 3, 6, 7

[23] Jian Liang, Dapeng Hu, Jiashi Feng, and Ran He. Umad: Universal model adaptation under domain and category shift. *arXiv preprint arXiv:2112.08553*, 2021. 2, 3, 6, 7

[24] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE TPAMI*, 2021. 3

[25] Jian Liang, Yunbo Wang, Dapeng Hu, Ran He, and Jiashi Feng. A balanced and uncertainty-aware approach for partial domain adaptation. In *ECCV*, 2020. 7

[26] Zhengfa Liu, Guang Chen, Zhijun Li, Yu Kang, Sanqing Qu, and Changjun Jiang. Psdc: A prototype-based shared-dummy classifier model for open-set domain adaptation. *IEEE Transactions on Cybernetics*, 2022. 1, 2

[27] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, 2018. 2

[28] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. Transfer feature learning with joint distribution adaptation. In *ICCV*, 2013. 1, 2

[29] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, 1967. 4

[30] Zak Murez, Soheil Kolouri, David Kriegman, Ravi Ramamoorthi, and Kyungnam Kim. Image to image translation for domain adaptation. In *CVPR*, 2018. 2

[31] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *CVPR*, 2019. 3

[32] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *ICCV*, 2017. 1, 2

[33] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019. 2, 6

[34] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 2, 6

[35] Sanqing Qu, Guang Chen, Jing Zhang, Zhijun Li, Wei He, and Dacheng Tao. Bmd: A general class-balanced multi-

centric dynamic prototype strategy for source-free domain adaptation. In *ECCV*. Springer, 2022. 1, 2

[36] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987. 2, 3, 4, 5

[37] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 2, 5, 6

[38] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. In *NeurIPS*, 2020. 1, 2, 7

[39] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. In *ICCV*, 2021. 2, 6, 7

[40] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018. 1, 2

[41] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *ECCV*, 2018. 2, 7

[42] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 5

[43] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001. 4

[44] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017. 2, 5, 6

[45] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555, 2017. 1, 2

[46] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *ICCV*, 2021. 1, 2

[47] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *CVPR*, 2019. 1, 2, 6